복수 모델영상의 상위도 패턴을 이용한 손형상 인식

이 칠 우[†]

요 약

본 논문에서는 복수 모델영상 간의 휘도치의 차이(이를 상위도라고 정의함)를 기본으로 영상매칭의 기준이 되는 「상위도 패턴」을 만드는 방법과, 이를 이용하여 입력영상과 모델영상간의 거리 계산에 의해 손의 형상을 인식하는 알고리듬에 대해 기술한다. 이 방법의 특징은 인식과정에서 에지나 코너 등과 같은 기하학적 형상특징을 이용하는 것이 아니고, 영상전체를 하나의 처리단위로 하여 다수의 영상으로부터 구한 단순한 휘도치 차이를 2차원의 형상으로 패턴화하여 인식과정에 사용한다는 점이다. 따라서 대량의 모델이 상호간에 갖는 상관관계가 인식과정에서 동시에 적용되기 때문에 대규모의 영상을 비교할 필요가 있는 인식 알고리즘의 실현에 매우 적합하다.

Hand Shape Recognition with Disparity Pattern of Multiple Model Images

Chil-Woo Lee[†]

ABSTRACT

This paper describes a method for making the "disparity pattern" which is basis of image matching with brightness difference; called disparity, between multiple model images, and an algorithm which recognizes hand shape by utilizing the pattern in measuring the distance between a input image and model images. The virtue of the algorithm is that only simple brightness difference calculated from multiple images by managing a whole image as the fundamental processing unit is patterned in two dimensional shape and then is used in the recognition process. Consequently, this method is very useful for other recognition algorithm requiring comparison of large scale image since correlation among multiple model images is applied simultaneously in recognition process.

1. 서 론

일상생활에서 인간은 다양한 수단을 사용하여 상대방에게 자신의 의사를 전달한다. 이 경우 흔히 언어적 수단만을 고려하기 쉬우나 몸짓, 손짓, 얼굴의 표정 등과 같은 비언어적 수단이 보다 많은 정보를 전달한다는 연구결과가 보고되어 있다[1]. 최근에 들어 컴퓨터 기술의 급속한 발전과 함께 미래 정보화사회를 향한 고기능의 정보 시스템 구축을 위한 갖가지 연구가 진행중이나 인간과 의사소통이 자유스럽고 보다 지적인 시스템을 제작하기 위해서는 무엇보다도 이러한 비언어적 대화 수단을 자동화함이 필요하다.

그 중에서도 손은 인간에게 있어서 가장 중요한 도구의 하나로 그 동작과 모양이 의미하는 바를 인식한다는 것은 정보 시스템과 인간 사이의 자연스러운 인터페이스를 실현할 수 있다는 점에서 오랜 기간동안 많은 사람들의 연구테마로 다루어져 왔다[10]. 그러나 인간의 손은 대단히 복잡한 관절구조를 가지고 있을 뿐만 아니라 인체에서 움직이는 속도가 가장 빠르고 그 형상이 순간적으로 변하기 때문에 이를 인식하여 입력수단으로 시스템에 직접 활용하기는 매우 어렵다. 특히 시각정보만을 이용하여 본 논문에

본 연구는 '95년도 전남대 학술연구비의 지원을 받아 수행되었음.

[†] 전남대학교 컴퓨터공학과

서와 같이 손 형상으로부터 문자를 인식한다는 것은 매우 곤란한 문제이다.

손 형상 인식의 자동화에 관한 연구는 크게 물리적인 센서 정보를 이용하는 방법과 비디오 카메라를 통하여 얻어지는 영상을 이용하는 방법으로 나눌 수 있다. 물리적인 센서를 이용하는 방법 중 가장 일반적인 것은 형상입력장치인 Data Glove를 사용하는 것이다[2]. 이 방법은 손가락을 구부릴 때 굴절에 의해 변화하는 빛의 양을 감지하여 구부린 각도를 측정하는 것으로, 측정이 쉽고 자기 센서와 조합하여 사용하면 손의 방향 및 대략적인 3차원 위치를 알 수 있으므로 현재 가상현실 시스템과 같은 인터렉티브 정보시스템의 입력장치로 가장 많이 쓰이고 있다. 그러나 이 방법은 손가락의 자세한 형상을 인식하기 위해서는 아직 정밀도가 낮고 전체 시스템과 접속하기 위해 반드시 연결선이 필요하다는 점에서 자연스런 인터페이스 구축에는 많은 문제점을 안고 있다.

손에 별도의 센서를 장착하지 않고 손 형상을 인식하기 위해서는 시각적인 방법을 이용하는 수밖에 없다. 그러나 시각적인 방법은 모델구축의 어려움과 세그멘테이션의 어려움 때문에 알기 쉬운 표식을 붙인다거나[3] 손의 형상이 뚜렷이 분리될 수 있는 간단한 문제를 위주로 적용되어 왔다. 또 난이도와 해상도의 문제 때문에 손가락 의 형상의 자세한 인식보다는 상반신을 포함한 어깨 전체의 제스처인식을 연구 목표로 삼고 있는 경우가 많다[10].

손형상 인식의 대표적인 방법으로 화소간의 휘도경도를 구해 히스토그램을 만들고 이 히스토그램을 평활화하여 얻어진 특징 벡터를 형상 구분의 척도로 삼는 방법이 있다[4]. 즉 모델영상과 테스트 영상으로부터 얻어진 특징 벡터간의 거리를 측정함에 의해 인식할 수 있게 된다. 특징 벡터간의 거리에 의한 모델영상을 평면에 나열하고 인식과정에서 거리를 보간함으로써 두 모델의 중간정도의 형상인식도 가능하다. 또, 'Size Function'이라는 특수한 방법으로 추출한 특징벡터를 이용하여 손모양으로 표현하는 문자를 인식하는 방법도 널리 알려져 있다[5].

손형상 연구와 관련된 최근의 연구성과로 그림자를 이용하는 방법[11]과 복수개의 카메라를 사용하는 방법[12]을 들 수 있다. 전자의 경우 간단한 손가락 형상을 대상으로 손영역과 손에 의해 생긴 그림자의 국소특징정보(peak와 valley)를 이용하여 3개의 손형상과 관련된 제스처를 인식할 수 있다. 그러나 이 방법은 손과 그림자 투영면과의 거리가 달라지면 특징정보의 매칭이 불안해 안정성에 문제점을 안고 있다. 후자의 경우는 여러 개의 카메라를 사용하여 인식에 필요한 최적의 영상을 선택한 다음 손영역의 중심점과 손가락의 주방향을 계산하여 8개의 명령을 구별할 수 있는 시스템을 제작하였다. 그러나 이 방법은 손가락의 끝부분이 항상 영상의 윗부분(즉 가상현실 시스템의 디스플레이 방향)을 향하고 있어 환경의 제약을 받는다는 점과 여러 개의 카메라를 사용한다는 점에서 시스템이 커지고 계산량이 많아진다는 단점이 있다.

카메라를 통하여 들어 온 영상만을 근거로 손의 형상을 인식하는 데는 크게 두 가지의 문제점이 따른다. 첫 째는 손은 인체 중에서 가장 많은 자유도를 지닌 부위로 평상시 각 구성체인 손가락이 서로 겹치거나 보는 시각에 따라 거의 무한종류의 형상을 지니고 있어 인식을 위해 필수적인 세그멘테이션이 매우 어렵다는 점이다. 즉 손의 특정 형상에 안정적으로 대응되는 에지나 코너 등의 기하학적 특징 추출이 어렵게 된다. 이 때문에 정확한 형상을 얻기 위해서는 불편하더라도 물리적인 센서를 직접 부착을 하거나, 손가락간에 서로 겹치는 현상이 일어나지 않는 제한적인 모양만을 인식의 대상으로 삼는 수밖에 없다.

둘째는 손은 크기에 비해 움직이는 범위가 매우 넓어 적절한 해상도의 영상을 얻기가 어렵다는 점이다. 즉, 형상을 인식하기 위해서는 카메라를 매우 가까이 접근시켜서 촬영하거나 렌즈를 사용하여 충분한 크기로 영상을 확대할 필요가 있는데 그 크기에 비해 상대적으로 손이 움직이는 속도가 매우 빠르고 범위가 넓기 때문에 움직임 정보를 유효하게 이용하기가 어렵게 된다. 따라서 수화인식과 같은 실제 응용에서는 원거리의 카메라를 이용하여 손의 대략적인 위치를 알아내는 과정과 손의 위치에 초점거리가 짧은 카메라의 시선을 맞추어 그 형상을 인식하는, 즉 해상도가 다른 두 가지 인식과정이 필요하게 된다.

본 논문은 상위도라고 정의한 입력영상과 모델영상과의 휘도치 차만을 이용하여 손의 형상을 인식하는 연구에 대해서 기술한다. 휘도치의 차이란 화소간의 뺄셈에 의해여 구해지므로 결국 두 영상간의 상관관계를 의미하나 2.1절의 식(1)에 의해 구해지는 값은 두 영상의 서로 다른 정도를 나타내므로「상위도」

라는 용어를 특별히 정의하여 사용하기로 한다.

본 논문에서 채택한 인식방법은 여러 가지 손 형상에 대한 모델영상과 참조영상을 미리 준비하고, 이들 사이의 상위도를 구하여 그 결과를 2차원의 패턴으로 형상화한 후, 이를 이용하여 입력영상의 각 모델에 대한 거리를 계산함으로써 형상을 인식하게 된다. 본 연구의 특징은 종래 영상인식에 있어서 중요시되었던 에지나 코너 등과 같은 형상특징정보를 이용한 것이 아니고, 다수의 모델영상사이의 상위도 만을 기초로 하여 "인식한다" 라는 입장에서 손의 형상을 신속히 분류해 낼 수 있다는 것이다. 따라서 이 방법은 최근 주목을 받고 있는 "메모리 기반 인식[6]" 과 "외관기반 인식[5,9]" 과 같이 특징 추출과정에서 필요한 각종 파라메터 설정문제를 피할 수 있어 실세계 시스템의 구현이 매우 수월해진다. 아울러 손과 같이 형상변화가 복잡하여 인식이 매우 어려운 물체를 영상간의 연산만으로 구할 수 있다는 새로운 사실을 보여주는 구체적인 예이다.

이 절 이후의 논문의 내용은 다음과 같다. 2절에서는 미리 준비한 영상들 사이에얻어진 상위도를 제약으로 2차원 영상배열을 작성하고 이를 형상화하여 인식의 기준이 되는 상위도 패턴을 만드는 방법에 대해서 설명한다. 3절에서는 상위도 패턴을 기본으로 하여 입력영상과 모델영상들의 거리를 측정하여 형상을 인식하는 방법에 대해서 기술한다. 4절에서는 제안하는 알고리즘을 고속으로 실현할 수 있는 하드웨어와 이를 이용한 실험결과에 대해서 기술하고 5절의 결론에서는 문제점 분석과 이후의 연구방향 등에 대해 기술한다.

2. 상위도 공간의 작성

2.1 상위도의 정의

본 논문에서 인식의 지표가 되는 영상들 사이의 상위도를 식(1)과 같이 정의한다.

$$D_{i,j} = Min_{(p,q)\in R_i}\{ \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} | I_i(p+m, q+n) - I_j(m, n) | \} \quad (1)$$

위 식이 나타내는 의미는 임의의 영상 I_i를 다른 영상 I_j와 영상영역 R_j내에서 가장 닮은 위치인 좌표 (p,q)에 겹쳤을 때 서로 겹쳐진 화소가 갖는 휘도치 차이의 절대값을 겹쳐진 영역($N \times M$ 범위)에 걸쳐 구한 최소 값이다. 이 $D_{i,j}$의 값은 화소간의 뺄셈에 의해 구해지므로 넓은 의미에서 두 영상의 상관관계(특히 영상이 급격히 변하지 않는 경우)를 나타낸다고 볼 수 있으나 엄밀히 보아 단순히 두 영상 I_i와 I_j가 서로 산술적으로 얼마나 다른가를 나타내므로 「상위도」라는 용어를 사용하여 표현하기로 한다.

2.2 모델영상과 참조영상의 수집

인간의 손은 여러 가지 복잡한 형상을 가질 수 있지만 본 논문에서는 25종류의 영문 알파벳을 나타내는 손가락문자를 인식 대상으로 하였다. 이후에 기술하는 알고리듬을 실현하기 위하여 미리 모델영상과 참조영상이라는 2개의 영상집단을 취득할 필요가 있다.

모델영상은 그림1과 같이 영문 알파벳(a,b,...y)을 나타내기 위해 사용되는 영상집합으로 손가락이 서로 겹치지 않고 가장 정확하게 인식할 수 있도록 손목이 회전하지 않는 상태로 취득한 영상을 말한다.

이에 비해 참조영상이란 인식의 기준이 되는 상위도 공간과 상위도패턴을 형성하기 위해 사용되는 영상집합으로 영상을 입력할 때의 형상의 변형 가능성을 고려하여 모델영상이 시선축을 기준으로 일정 범위

그림 1. American Sign Language의 표준으로 사용되는 알파벳 손가락문자(a-y) 영상. 왼쪽 위에서부터 순서대로 a, b, c,⋯,y.

내에서 회전되고 자유스런 움직임으로 인해 약간의 형상변형이 가미된 영상을 말한다. 즉 모델영상에 다소의 잡음을 가한 영상이라고 볼 수 있다.

참조영상은 손과 카메라가 일정거리를 유지하고 손목을 광축을 기준으로 -45°에서 +45°의 범위에서 자연스럽게 회전시키면서 입력시킨다. 이 때 전 영상을 참조영상으로 간주하는 것이 아니고 새로 취득할 영상과 이미 취득된 모든 영상과의 상위도를 계산하여 일정 값 이상이 될 때 그 영상을 새로운 참조영상으로 등록한다. 이 방법은 인위적인 관찰에 의해 형상을 판단하기보다는 인식에 필요한 기준에 입각하여 시스템이 영상데이터를 자연스럽게 획득할 수 있다는 점에서 많은 응용이 기대된다. 참조영상의 일부를 그림2에 보인다.

2.3 상위도 공간의 생성

영상간의 상위도, 즉 영상이 지닌 밝기의 산술적 차이만을 비교하여 손과 같이 복잡한 물체의 형상을 인식한다는 것은 매우 어려운 일이다. 입력영상과 시스템이 지니고 있는 모델영상을 하나 하나 비교하여 가장 유사한(즉 상위도가 가장 적은) 영상을 대응하는 손 형상으로 분류할 수는 있으나 영상을 입력할 때 손목의 회전과 손가락의 겹침, 크기의 변화 때문에 안정된 결과를 얻는다고는 보장하기 어렵다. 따라서 본 논문에서는 입력영상과 수많은 참조영상(본 논문의 경우 256개)을 인식과정에서 한꺼번에 비교하여 안정된 결과를 얻을 수 있는 방법을 제안한다.

상위도공간이란 상위도를 척도로 입력영상의 위치를 규정하는 2차원 직교공간을 말한다. 이 공간내에서 각 영상들은 상호간에 상위도 만큼의 거리를 두고 떨어져 놓이게 된다. 따라서 서로 비슷한 영상은 상위도가 적으므로 근접한 위치에 놓여 하나의 집단을 이루게 된다.

그림3에는 4매의 영상($I_1 \sim I_4$)이 순서대로 이 공간에 놓여지는 모습을 보여주고 있다. 먼저 영상 I_1를 이 공간의 원점 ($x_1=0, y_1=0$)에 놓고 다음 영상 I_2를 X축 선상이면서 두 영상간의 상위도 $D_{1,2}$만큼 떨어진 위치($x_2, y_2=0$)에 놓는다. 이를 기준으로 두 영상과 I_3간의 상위도 $D_{1,3}$와 $D_{2,3}$를 만족하는 위치 (x_3, y_3)에 I_3를 배치할 수 있다. 3장의 영상만을 고
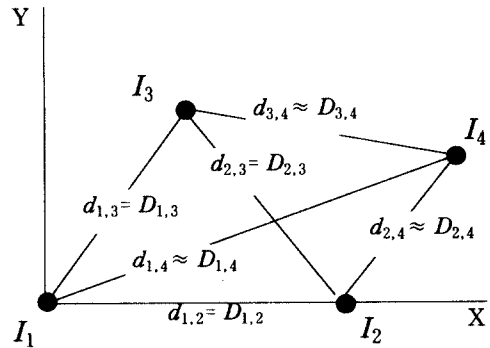
그림 3. 상위도공간에서 4매의 영상($I_1 \sim I_4$)을 배치시킨 모습(본문참조). "≈"기호는 두 영상간의 상위도 $D_{i,j}$와 공간상의 거리 $d_{i,j}$가 근사적으로 일치함을 의미한다.

그림 2. 참조영상(1300개로 구성된 전체영상에서 일부분만을 잘라서 표시한 부분영상). 카메라 시선축을 기준으로 손목을 -45° 방향에서 +45° 방향으로 서서히 회전시키면서 앞서 취득된 영상들과 현재 입력영상 간의 상위도가 일정 값 이상이 될 때 입력영상을 새로운 참조영상으로 등록한다. 한 문자에 대응하는 손영상이 각도를 약간씩 달리하면서 입력되어 있음을 알 수 있다.

려할 경우 위 방법에 의해 유일한 위치를 결정할 수 있다. 그러나 네 번째 영상 I_4를 추가로 배치할 경우 기존의 세 영상과 I_4간의 상위도를 이용하여 2차원 공간상에서 상호간의 제약을 모두 만족하는 위치를 구할 수는 없다. 이 경우 고차원으로 문제의 공간을 확대하면 해를 구할 수 있으나(예를 들어 삼각뿔 형태의 구조) 문제를 간단히 하기 위해 2차원 공간만을 고려할 경우 그림3처럼 복수의 상위도를 근사적으로 만족시키는 위치에 I_4를 배치시키는 것이 최선의 방법이 된다. 이를 확장하여 임의 개수의 영상을 2차원 공간에 배치하는 방법을 설명하면 아래와 같다.

하나의 참조영상이 2차원 공간의 임의의 위치에 놓여있고 이 영상과 다른 모든 영상이 상위도 만큼의 길이를 갖는 용수철로 서로 연결되어 있다고 가정하자. 자연상태의 경우 용수철간의 탄력에 의해 영상들의 위치가 자연스럽게 결정되게 되나 수학적으로는 상위도를 제약으로 하는 연립방정식을 풀면 영상의 위치를 구할 수 있다. 그러나 상위도란 휘도치의 단순한 산술적인 차이이므로 위에서 설명한 바와 같이 복수의 영상을 구별하는 선형적인 척도로 사용될 수 없어 연립방정식의 근을 구할 수 없다. 따라서 주어진 제약을 최대한으로 만족시키는 최적해를 구할 수 밖에 없다. 이것은 모든 용수철이 가지고 있는 에너지의 합계가 최소가 되는 점들의 위치를 구하는 것을 의미한다.

평면에 배치한 N개의 영상을 I_i $(1 \le i \le N)$이라고 하고 이 영상이 위치한 2차원 평면의 좌표를 (x_i, y_i)라고 하자. 평면 위의 두 영상 $P_i(x_i, y_i)$와 $P_j(x_j, y_j)$사이의 거리는 다음 식으로 구해질 수 있다.

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

최소 에너지를 갖게 하기 위해서는 식(1)에 의해 구한 두 영상간의 상위도 $D_{i,j}$와 평면상의 거리 $d_{i,j}$가 일치하도록 영상을 배치하면 된다(이 때 용수철이 갖는 에너지는 영이 된다). 그러나 복수의 영상이 한 점에 연결되어 있고 상위도는 산술적으로 선형이 아니기 때문에 힘의 불균형이 발생하여 용수철의 에너지는 절대로 영이 되지 않는다. 최적의 근사해를 구하기 위해서는 식(3)으로 표현되는 에너지의 합 E_{sum}이 최소가 되는 위치 (x,y)를 식(4),(5)를 풀어 구하면 되지만 이 방법은 앞서 설명한 바와 같이 해가 수렴하지 않을 우려가 있으므로 간단한 반복법에 의해 새로운 위치를 구한다. 즉 현재의 위치에서 식(6),(7)에 의해 구한 미소 변위 만큼 이동시켜가다가 모든 점들이 더 이상 움직이지 않을 때 계산을 멈추고 배치작업을 완료한다.

$$E_{sum} = \sum_{i=1}^{N} \sum_{j=1}^{N} (D_{ij} - d_{ij})^2 \quad (3)$$

$$\frac{\partial}{\partial x} E_{sum} = \frac{\partial}{\partial x} \sum_{i=1}^{N} \sum_{j=1}^{N} (D_{ij} - d_{ij})^2 = 0 \quad (4)$$

$$\frac{\partial}{\partial y} E_{sum} = \frac{\partial}{\partial y} \sum_{i=1}^{N} \sum_{j=1}^{N} (D_{ij} - d_{ij})^2 = 0 \quad (5)$$

$$\triangle x = \frac{1}{N} \sum_{j} \frac{x_i - x_j}{D_{ij}} (D_{ij} - d_{ij}) \quad (6)$$

$$\triangle y = \frac{1}{N} \sum_{j} \frac{y_i - y_j}{D_{ij}} (D_{ij} - d_{ij}) \quad (7)$$

이상에서 설명한 방법을 참조영상에 가하여 새로운 위치를 구해보면 그림4와 같다. 그림에서 보는 바와 같이 영상들은 어떤 의미를 지닌 형태로 배치된다. 즉 서로 유사한 영상들은 가까운 곳으로 모이고 다른 영상들은 먼 거리를 두고 떨어지게 된다. 또 동일한 형상은 입력시의 손목의 회전 각도에 비례하여 원호를 그리면서 연속적으로 분포하게 된다. 이 결과는 상위도를 척도로 할 때 복수의 영상들간의 위상관계를 나타내는 것으로 이를 패턴화하여 인식에 사용할 경우 매우 안정된 결과을 얻을 수 있다.

그림4의 배열은 이산적(discrete)인 좌표값을 가지고 있지 않기 때문에 컴퓨터로 다루기가 매우 어렵다. 따라서, 그림4의 결과를 $n \times n$(본 논문에서는 n=16)
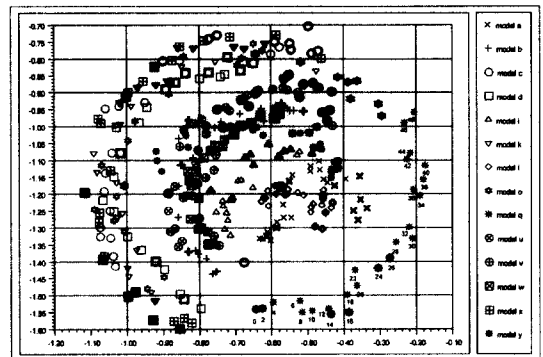
그림 4. 영상들간의 상위도를 제약으로 식(6)의 반복법에 의해 결정된 영상들의 평면배치. 유사한 영상은 서로 근거리에 위치하고 동일 형상을 지니면서 회전각도가 다른 영상은 동일 원주상에 존재한다.

의 영역으로 나누고 각각의 영역을 대표하는 영상을 선택하여 이산적인 형태로 재배치한다. 즉, 격자점 $G_i(1 \leq i \leq n \times n)$에 대해서 최근방에 있는 참조영상 S_j를 선택하여 이 영상의 새로운 위치로 설정한다. 이 공간을 이산상위도공간이라고 한다. 그림5는 이산상위도공간에 배치된 참조영상의 일부분을 보여준다.

그림 5. 이산상위도공간. 참조영상을 상위도를 근거로 배치하고 각 영상의 해당 위치를 16×16의 격자형태로 표현한 공간(본문참조).

3. 상위도 패턴을 이용한 형상인식

3.1 상위도패턴의 생성

상위도 패턴이란 2.3절에서 서술한 이산상위도 공간의 한 점 (p, q)에 배치된 참조영상과 임의의 영상과의 상위도를 구해 그 값을 0에서 255 사이의 값으로 정규화한 후 그 결과를 각각의 참조영상의 위치 (p, q)에 기록한 정방형의 패턴을 말한다. 손가락 문자를 나타내는 모델영상을 입력으로 하여 이 패턴을 만들면 하나의 모델영상에 대해 복수의 참조영상이 갖는 상관관계를 한꺼번에 나타낼 수 있다. 상위도 패턴의 생성과정을 상세히 기술하면 아래와 같다.

상위도공간의 한 점 (p, q)에 배치된 참조영상을 $R_{p,q}$라고 하자. 이 $R_{p,q}$와 임의의 영상과의 상위도를 구해 이를 $D_{p,q}$라 하고 그 값을 격자좌표 (p, q)에 기록한다. 이 값은 이산상위도공간이 이상적인 경우, 즉 참조영상의 수가 무한히 많고 형상변화가 서서히 일어날 경우는 급격한 변화 없이 연속적으로 변화하여야 하지만, 실제로는 상위도가 전혀 예측 불가능한 값을 지니므로 연속성이 보장되지 않는다. 이 연속성은 손형상의 미묘한 변화에도 안정된 인식결과를 얻을 수 있는가 하는 문제와 매우 밀접한 관계를 갖게 된다. 따라서 매끄러운 변화를 보증하는 범위의 상위도를 이용하기 위해서는 상위도가 어느 문턱치 th를 초과한 경우 그 상위도를 일정치에 설정하여 전체조합에 영향을 미치지 않게 함이 필요하다. 본 논문에서는 식(8)의 비선형변환을 이용하여 매칭과정에서의 안정화를 기한다.

$$D_{p,q} = \begin{cases} \alpha \times \sqrt{D_{p,q}} & \text{if } D_{p,q} < th \\ \alpha \times \sqrt{th} & otherwise \end{cases} \quad (8)$$

위 식에서 α는 $D_{p,q}$의 최대값이 255가 되게하는 임의의 상수이다.

비선형 변환을 실시한 후 주변 요소들과의 연속성을 확보하기 위하여 다시 3×3의 필터를 이용하여 평활화를 실시한다.

그림 6에 이상의 처리과정을 통하여 얻어진 한 상위도 패턴의 한 예를 보여준다. 이 패턴은 특정 모델영상과 상위도공간에 존재하는 참조영상과의 상관관계를 한꺼번에 나타내는 패턴이다. 따라서 이 패턴을 이용하면 매칭과정에서 다수(본 논문에서는 256

그림 6. 상위도 패턴의 예. 이산상위도공간 상의 임의의 한 영상과 특정영상과의 상관관계를 나타내는 패턴 (1-16의 숫자는 이산상위도 공간에서의 격자번호를 나타냄)

개)의 영상이 한꺼번에 사용될 수 있어서 매우 안정된 인식 결과를 얻을 수 있다.

3.3 상위도 패턴을 이용한 인식

그림7에 상위도 패턴을 이용한 인식처리의 전 과정을 보인다. 그림에서 알 수 있는 바와 같이 인식과정은 1)상위도 공간을 만드는 과정, 2)모델영상에 대한 상위도 패턴을 만드는 과정, 3)입력영상에 대한 상위도 패턴을 만드는 과정, 4)모델영상과 입력영상의 상위도 패턴의 비교에 의한 인식과정으로 크게 나누어진다. 1)과 2)과정은 시스템의 작동에 앞서 미리 계산해 놓고 3)과 4)과정은 전용의 하드웨어를 이용하면 신속히 계산해 낼 수 있다. 인식은 4)과정에서 손의 형상을 정확히 표현하는 모델영상의 상위도 패턴과 입력영상의 상위도 패턴 사이의 차이를 계산하여 최소 값을 나타내는 모델을 입력영상의 손 형상으로 인식한다.

4. 실험 및 고찰

4.1 손가락 문자의 모델영상

손의 위쪽 방향에 위치한 비디오 카메라를 통하여 들어오는 256계조의 흑백영상을 16×16 화소로 축소하여 입력영상으로 사용하였다. 입력영상은 그림2의 참조영상을 만들 때 의도적으로 가했던 회전의 범위를 넘지 않는 선에서 임의의 방향을 가지게 하여 촬영한다. 따라서 그림2의 참조영상에 나타나는 영상들의 연속적인 집합이 입력영상이 되게 된다. 이러한 영상의 경우 크기가 작을 때에는 'a'와 's', 'd'와 'r'은 상위도 만으로는 거의 식별이 불가능 할 정도로 유사하다. 그런 까닭에 손가락문자의 영상을 인간이 보아서 식별가능한 14종류의 손가락문자(a, b, c, d, i, k, l, o, q, u v, w, x, y)로 다시 분류하여 인식실험을 실시하였다.

하드웨어 시스템은 그림8에서 보여주는 것처럼 VME버스를 통해서 서로 접속 되어 있는 특수 영상처리장치(ReMOT-M 과 Max Video)와 호스트 컴퓨터인 SUN웍스테이션으로 구성되어 있다. ReMOT-M은 실험실에서 직접 제작한 것으로 상위도 연산을 고속으로 실행하는 것 이외에 영상의 이치화, 에지추출, 평활화등의 전처리를 실시간으로 실행할 수 있다[8]. 또, 상위도 연산 전용의 LSI를 4개 탑재하고 있으므로 동시에 4개의 영상에 대한 상위도 계산이 가능하다. 본 시스템에서는 Max Video에서 영상의 전처리를, ReMOT-M에서는 상위도연산을 실시간

그림7. 상위도패턴을 이용한 인식과정의 흐름도.

그림 8. 실험에 사용한 하드웨어의 구성

에 실시하고, 연산결과만을 호스트 컴퓨터에 전송하여 인식을 행한다.

700개의 손의 모델영상을 "평면배치 알고리즘"을 이용해서 2차원의 평면에 배치하고 그 중 350개만을 표시한 모양을 그림4에 나타내었다. 그림에서는 한 카테고리에 속한 50개의 모델영상의 위치가 원을 그린 것처럼 분산되어 있음을 잘 알 수 있다. 모델영상의 획득은 앞서 기술한 것처럼 카메라 광축에 수직인 평면 위에서 -45°에서 +45°의 범위까지 손을 회전시키면서 차례대로 입력하였는데, 이 그림에서 모델번호와 입력시의 각도변화가 잘 대응하고 있음을 알 수 있다. 즉 한 문자의 영상이 하나의 원의 궤적을 따라 질서 있게 배치되어 상위도공간은 모델영상의 각도변화의 특징을 나타내고 있음을 의미한다.

다른 모델도 동일한 모양으로 분포하고 있으므로 이 상위도공간에서의 하나의 카테고리에 대한 평면배치의 분산은 모델영상의 각도변화에 의해 생기는 것임을 알 수 있다. 또, 두 개의 카테고리의 모델이 같은 위치에 중복되는 경우가 있는데, 이것은 두 카테고리에 속한 모델영상이 매우 유사한 모양을 가지고 있어 상위도만을 척도로 해서는 서로 구별할 수 없기 때문에 일어난 결과이다. 이 점은 오류를 일으키는 원인으로 작용하게 되는데 상기한 바와 같이 상위도패턴을 이용하여 대규모 모델간의 매칭을 동시에 수행하게 됨으로써 인식이 안정되게 된다.

상위도패턴을 이용한 인식결과를 표1에 나타내었다. 오류는 모델 '*c*', '*d*'와 모델 '*o*'와 같이 상위도 공간 내에서는 거의 동일한 장소에 모델이 위치하고 있을 때 발생한다. 이들 세 모델은 그림4에서 알 수 있는 것처럼 참조영상의 상위도공간에서 매우 근접해 있

표 1. 상위도 공간을 이용한 인식결과(인식오류난에는 제일 인식오류가 많았던 문자의 비율을 표시하였음).

손가락문자	인식성공률 (%)	인식오류 (%)
a	100	-
b	84.0	u (10.0)
c	88.0	y (8.0)
d	84.0	c (6.0)
i	100	-
k	98.0	y (2.0)
l	98.0	y (2.0)
o	92.0	i (6.0)
q	84.0	l (6.0)
u	88.0	y (8.0))
v	84.0	w (6.0)
w	96.0	u (4.0)
x	86.0	o (8.0)
y	96.0	l (4.0)

다. 따라서 서로 매우 유사한 상위도패턴을 지니게 되어 구별이 어렵게된다.

이 문제를 해결하기 위해서는 보다 세밀한 레벨의 상위도공간을 작성할 필요가 있다. 상위도공간을 계층성을 지닌 부분공간에 분할하여, 해상도의 낮은 공간에서 차례로 식별해 가면 인식은 가능하리라고 생각되어진다. 해상도가 낮은 공간에서 식별할 수 없었던 유사모델도 보다 해상도가 높은 부분공간에서는 식별할 수 있는 가능성이 있다.

5. 결 론

본 논문에서는 두 영상의 상위도를 기반으로 하여 손가락에 의해 표현되는 영어문자를 안정되게 인식할 수 있는 알고리즘에 대해 기술하였다. 제안한 방법의 특징은 입력영상의 특징정보를 인식의 수단으로 삼지 않고, 다수 모델영상의 휘도정보를 그대로 사용한 점이다. 이 방법을 사용하면 손과 같이 유연한 경계선을 갖는 물체의 특징(예를 들어 에지나 코너) 정보 추출시 필요한 파라메터 설정 문제를 피할 수 있으며 영상의 휘도치 정보를 그대로 이용하기 때문에 매우 안정된 인식결과를 얻는다.

실험에서 사용한 14 카테고리의 각 50장(합계700

장)의 모델영상을 표현하기 위해서 16×16화소의 "이산상위도공간"이 충분한가에 대해서는 확고한 검증이 필요하다. 따라서 다음의 연구에서는 "상위도공간"의 크기에 의한 식별능력이 어떻게 변하는가를 평가하는 점이 연구과제로 남는다. 또 특정 카테고리에 속하는 모델영상에서 얻어진 상위도공간 분포가 어느 정도의 변화까지 같은 카테고리로 식별할 수 있을까 하는 지표인 "허용범위"에 대해서도 평가해볼 계획이다.

본 논문에서 제안한 방법은 "계측"이 아닌 "인식"이라는 입장에서 한 장의 영상을 기본단위로 하여 종래의 연구에서 중시해온 2차원 특징 및 3차원 정보를 위한 어떤 중간표현도 없이 휘도치 만을 이용하여 손과 같이 복잡한 형상을 지닌 물체를 신속히 인식해 낼 수 있는 장점을 가지고 있다. 그러나 손가락의 세밀한 형상을 인식하는데는 어느 정도의 한계를 지닌다. 따라서 세밀한 형상인식이 필요한 경우 주어진 목표에 맞게 개략적인 형상인식을 통해 탐색영역을 좁힌 다음 동일한 과정을 상세 모델에 대해 연속적으로 적용함으로써 인식이 가능하게 된다. 제안하는 방법은 매우 규모가 큰 영상집합을 데이터로 사용하고 있다는 점에서 최근 주목을 받고 있는 "메모리 기반 추론(memory based reasoning)" 및 "외관기반 영상인식(appearance based image recognition)"에 대한 하나의 비교 대상이 될 수 있다.

참 고 문 헌

[1] Raffler-Engel(ed.), "Aspects of Nonverbal Communication", Swets and Zeitlinger B. V., 1980

[2] J. Ohya and Y. Kitamura, etc. "Real-Time Reproduction of 3D Human Images in Virtual Space Teleconferencing," in Proc. of '93 IEEE Virtual Reality Annual Int. Symp. pp.408-414, 1993

[3] R. Cipolla, Y. Okamoto, and Y. Kuno, "Qualitative visual interpretation of 3D hand gestures using motion parallax," in Proc. of MVA'92, pp.477-482, 1992

[4] Willliam T. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition," in Course Note of SIGGRAPH'94, pp.M-1-M-23, 1994

[5] C. Uras and A. Verri, "On the Recognition of the Alphabet of the Sign Language Through Size Function," in Proc. of 12th IAPR, Vol. 2, pp.334-338, 1994

[6] H. Kitano, "A Comprehensive and Practical Model of Memory-Based Machine Translation," in Proc. of IJCAI'93, pp.1276-1282, 1993

[7] C. W. Lee, A. Tsukamoto, K. Hirota, and S. Tsuji, "A Visual Interaction System Using Real-Time Face tracking," in Proc. of 28th Asilomar Conference on Signals, systems & Computers, pp.1282-1286, 1994

[9] Hiroshi Murse and Shreek. Nayar, "Visual Learning and Recognition of 3-D Objets from Appearance", int. Journal of Computer Vision, Vol.14, 1995

[10] V. Pavlovic, R. Sharma, and T. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Trans. on PAMI, Vol.19, No.7, pp.677-695, 1997

[11] J. Segen and S. Kumar, "Shadow Gesture: 3D hand Pose Estimation using a Single Camera", in Proc. of CVPR'99, pp.479-485, Colorado, 1999

[12] A. Utsumi and J. Ohya, "Multiple Hand Gesture Tracking using Multiple Cameras", in Proc. of CVPR'99, pp.473-478, Colorado, 1999

이 칠 우

1986년 중앙대학교 전자공학과 졸업
1988년 중앙대학교 대학원 전자공학과 졸업(공학석사)
1992년 동경대학 대학원 전자공학과 졸업(공학박사)
1992년~1995년 이미지 정보과학연구소 수석연구원 겸 오사까대학 기초공학부 협력연구원
1995년 리츠메이칸대학 특별초빙강사
1996년~현재 전남대학교 공과대학 컴퓨터공학과 교수
관심분야 : 컴퓨터비전, 컴퓨터그래픽스, 멀티미디어 데이터베이스